ISNI and VIAF: authority files and identity management

Andrew MacEwan, British Library & ISNI International Agency
VIAF and ISNI: scope

**VIAF Scope**
- Persons
- Organisations
- Works / uniform titles
- Expressions
- Meetings
- Geographic
- All public data

**ISNI Scope**
- Persons
  - + musicians, researchers
- Organisations
  - (excluding sparse)
  - (excluding undifferentiated)
- Includes private data
VIAF and ISNI: purpose

**VIAF Role**
- Ingest authority records from the world’s major national and research libraries
- Make clusters
- Expose and diffuse

**ISNI Role**
- Create permanent IDs
  - By batch
  - On demand
- Diffuse those IDs
  - Library Services, e-commerce, Rights Management, Research Management, Publishing, Linked Open Data, etc…
### VIAF System
- Harvester
- Clustering mechanism (re-clustered monthly)
- 5 web interface languages
- Download in multiple formats
- Linked data & SRU

- 1 million personal visitors p.a.

### ISNI System
- Batch load
- Online request API
- Web site (English only)
  - Allows end user input
  - Member input and correction
  - 16+ indexes
- SRU; linked data
- Quality Team monitoring & correcting
- Diffusion, including corrections
VIAF contributors...

- Library of Congress/NACO
  - National Library of Mexico
  - British Library
  - National Agricultural Library (U.S.)
  - National Library of Medicine (U.S.)
  - National Library of New Zealand
  - National Library of Scotland
  - National Library of South Africa
  - National Library of Wales

- German National Library
- National Library of France
- National Library of Sweden
- National Library of Australia
- National Library of Spain
- National Library of Portugal
- Central Institute for the Union Catalogue of the Italian libraries
- National Library of the Czech Republic
- National Library of Israel
  - Israel Museum
- Library of Alexandria, Egypt
- Vatican Library
- Swiss National Library
- Library and Archives Canada
- Union List of Artist Names [Getty Research Institute]
- NUKAT Center of Warsaw University Library
- National Széchényi Library, Hungary
- RERO - Library Network of Western Switzerland

- Sudoc [ABES], France
- Flemish Public Libraries
- National Library of Russia
- National Library of the Netherlands
- National Library of Norway
- DBC (Danish Bibliographic Center)
  - Danish Agency for Culture
- National Diet Library, Japan
- National Library Board, Singapore
- National Library of Latvia
- National Library of Poland
- National Library of Catalonia
- Lebanese National Library
- Perseus
- Syriac Reference Portal
- Wikidata
- ISNI
- National Library of Ireland
- National and University Library in Zagreb
- National Central Library, Taiwan
- National Library and Archives of Québec
- National Library of Korea
- National Library of Luxembourg
- National Library of Chile
- National Library of Morocco
## Current ISNI Sources …and growing

### GENERAL SOURCES

<table>
<thead>
<tr>
<th>Organization</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bowker Books in Print</td>
<td>BOWKER</td>
</tr>
<tr>
<td>The European Library (48 national libraries)</td>
<td>TEL</td>
</tr>
<tr>
<td>Virtual International Authority File (33 libraries)</td>
<td>VIAF</td>
</tr>
</tbody>
</table>

### RIGHTS MANAGEMENT

<table>
<thead>
<tr>
<th>Organization</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access Copyright, Canada</td>
<td>ACCE</td>
</tr>
<tr>
<td>Authors’ Licensing and Collecting Society, UK</td>
<td>ALCS</td>
</tr>
<tr>
<td>Centrum Dienstverlening Auteurs- en aanverwante Rechten, Netherlands</td>
<td>CEDA</td>
</tr>
<tr>
<td>Centro Español de Derechos Reprográficos</td>
<td>CEDR</td>
</tr>
<tr>
<td>Irish Copyright Licensing Agency</td>
<td>ICLA</td>
</tr>
<tr>
<td>Prolitteris, Switzerland</td>
<td>PROL</td>
</tr>
<tr>
<td>VG WORT, Germany</td>
<td>VGWO</td>
</tr>
</tbody>
</table>

### ORGANISATIONS

<table>
<thead>
<tr>
<th>Organization</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Chemical Society</td>
<td>ACS</td>
</tr>
<tr>
<td>Boekenbank, Belgium</td>
<td>BOEK</td>
</tr>
<tr>
<td>Bowker Publishers</td>
<td>BOWP</td>
</tr>
<tr>
<td>Publishers Licensing Society, UK</td>
<td>PLS</td>
</tr>
<tr>
<td>Ringgold</td>
<td>RING</td>
</tr>
</tbody>
</table>

### MUSIC

<table>
<thead>
<tr>
<th>Organization</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Musicological Society</td>
<td>AMS</td>
</tr>
<tr>
<td>British Library Sound Archive</td>
<td>BLSA</td>
</tr>
<tr>
<td>International Performers’ Database Association</td>
<td>IPDA</td>
</tr>
<tr>
<td>MusicBrainz</td>
<td>MUBZ</td>
</tr>
</tbody>
</table>

### RESEARCHERS AND PROFESSIONALS

<table>
<thead>
<tr>
<th>Organization</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Musicological Society</td>
<td>AMS</td>
</tr>
<tr>
<td>Authors Guild</td>
<td>AGLD</td>
</tr>
<tr>
<td>British Library Theses</td>
<td>BRTH</td>
</tr>
<tr>
<td>Digital Author identifier, Netherlands</td>
<td>DAI</td>
</tr>
<tr>
<td>Jisc Names Project, UK</td>
<td>JNAM</td>
</tr>
<tr>
<td>La Trobe University</td>
<td>AU:VLU</td>
</tr>
<tr>
<td>Modern Languages Association</td>
<td>MLA</td>
</tr>
<tr>
<td>OCLC Theses</td>
<td>OCLCT</td>
</tr>
<tr>
<td>ORCID and DataCite Interoperability Network</td>
<td>ODIN</td>
</tr>
<tr>
<td>AuthorClaim and RePec</td>
<td>OPENL</td>
</tr>
<tr>
<td>Proquest Theses</td>
<td>PROQ</td>
</tr>
<tr>
<td>Scholar Universe, Proquest</td>
<td>SCHU</td>
</tr>
<tr>
<td>Electronic tables of content</td>
<td>ZETO</td>
</tr>
</tbody>
</table>
Current ISNI network members … and growing

**Current ISNI-IA Members:**

- ABES (French Bibliographic Agency for Higher Education)
- CEDRO (RMO)
- CDR (Centrale Discotheek Rotterdam)
- Copyrus (RMO)
- FCCN (Research)
- French National Archives (Archives nationales de France)
- Harvard University
- Iconoclaste
- Irish Copyright Licensing Agency (ICLA) (RMO)
- ISSN International Centre
- La Trobe University
- Library of Congress
- National Library of Finland
- National Library of Norway
- National Library of New Zealand
- Publishers Licensing Society
- UNSW Library

**List of Registration Agencies:**

- **BnF (Bibliothèque nationale de France)** - primarily (though not exclusively) registers names of individuals and publishing organizations in France
- Bibliothèque nationale de Luxembourg
- British Library
- Electre
- Koninklijke Bibliotheek
- Kültür ve Turizm Bakanlığı
- National Assembly Library of Korea
- National Library of Korea
- National Library of Poland
- Numerical Gurus
- **Ringgold** - registers organization names
Synchronisation ISNI to VIAF

- **2012**: ISNI / VIAF identifiers
- **2013**: Full records; ISNI a VIAF source
- **2014**: ISNI records, verification mark
ISNI as Authority Data/
ISNI as Identifier Management system

• ISNI data sources
  • Assignment
  • Bridging data
  • Bib and authority data

• Searching and indexes

• Maintenance
  • Web maintenance

• Hunting anomalies

• VIAF and interoperability
  • Data characteristics
  • Movement reports
  • Cluster errors

Quality Team (& Members)
  • Curated crowd sourcing
  • Merging & splitting
  • Sampling
  • Discussion & policy papers
Identifier Management Ecosystem

- Algorithms
- Notifications
- Data fixing
- Sampling
- Data Policy
- Enrichment
- Correction
- Curation
- Crowd sourcing

ISNI Database
Harvested, Batch loaded; Online contributions

Ongoing data contribution – Members and Registration Agencies
# ISNI Statistics November 2014

- **Assigned:** 8.69 million ISNIs
  - **Provisional: Possible:** 700,815
  - **Provisional: Unassigned:** 9,287,278

## Assigned ISNIs November 2014

<table>
<thead>
<tr>
<th>Source Type</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIAF + non VIAF sources</td>
<td>4,870,099</td>
</tr>
<tr>
<td>3+ VIAF sources</td>
<td>428,988</td>
</tr>
<tr>
<td>2+ sources (not VIAF)</td>
<td>315,915</td>
</tr>
<tr>
<td>Unique name</td>
<td>2,735,449</td>
</tr>
<tr>
<td>Trusted single source (JISC, BOEK, RING)</td>
<td>342,231</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>8,692,683</td>
</tr>
</tbody>
</table>

- **Authoritative, Unique, Trustful, Persistent:** 8.24 million persons, 446,258 organisations
ISNI Assignment: Batchload/Matching criteria

Scores are collected from each judge (ice skating style)

Lowered for common surnames and common titles

Score > .85 = match

Score > .6 but < .85 = possible match

Name
Title
Partial title
Rare title word
Date
Publisher
Personal affiliation
Organisation affiliation
ISBN, ISWC, ISAN, DOI +
Other name identifier e.g. IPI, VIAF, IPD
Instrument
Linked entities
Dewey classification
Confidence

The two main problems for maintaining persistence are
- duplicates needing to be merged
- undifferentiated identities needing to be split

ISNI errs on the side of making duplicates rather than mixed identities

Thus the batch load process (usually) makes a provisional record
- where there is no match (for fear of making a duplicate assignment)
- where there is a low confidence match (for fear of making a mixed identity or a duplicate assignment)
- where a matching record already has another local ID for the same source, regardless of the strength of the match (for fear of making a mixed identity)
Possible Matches with low confidence are flagged

Possible merge: 125736770 92.842
Record no.: 12741357X
Name: Kovacs, Katalin
Creation role: author
Titles: Effects of platelet-rich plasma on the ossification induced by the synthetic
Notes:
Sources: PROQ

Click
<table>
<thead>
<tr>
<th>Possible merge: 125736770 92.842 Possible match under comparison →</th>
<th>Record no.: 125736770</th>
</tr>
</thead>
<tbody>
<tr>
<td>Record no.: 12741357X Name: Kovacs, Katalin</td>
<td></td>
</tr>
<tr>
<td>Creation role: author Titles: Comparative study of ossification induced by a synthetic bone-substitute material (beta-tricalcium phosphate) versus platelet-rich plasma in dogs</td>
<td></td>
</tr>
<tr>
<td>Titles: Effects of platelet-rich plasma on the ossification induced by the synthetic bone-substitute material, beta-tricalcium phosphate (comparative study on bone formation in dogs) Notes: Sources: PROQ</td>
<td></td>
</tr>
<tr>
<td>Notes: Sources: PROQ</td>
<td></td>
</tr>
</tbody>
</table>
Procedures for maximizing assignment

- Refinement of matching algorithms
  - E.g. introduced rare title word;
  - Now ignoring date of birth 1900
- Re-import program
  - Rematch with new rules
  - Rematch after new data added
- ISNI Quality Team: Data sampling
  - assessing impact of single source
  - Recommendations for program changes
- New criteria
  - Assessing uncommon surname assignment
  - Rules for online rich assignment
Assignment: Batch loading

Unique name

ISNI: 0000 0001 2454 6145
Name: Bothe, Jacob
Creation role: author

Notes:
Sources: VGWO
Publishers: Diogenes Verlag AG

ISNI: 0000 0004 0372 8224
Name: Centre Royal de Teledetection Spatiale
Name variant: CRTS
Royal Centre for Remote Sensing
Location / Nationality: Morocco Rabat
Creation class: Language material
Creation role: originator
Notes:
Sources: RING

Single source
Types of data batch loaded

- Authority Data
  - Most sources
  - Full evaluation
  - Normally during batch load if no match or not unique, loaded as provisional

- Bibliographic Data
  - E.g. Bowker, TEL, Zetoc
  - Forced where suspected name string authority control only
  - Only Name / title compared
  - Not admitted if no match, no possible match or not unique name
ISNI Members

• View whole database (but not restricted fields)
• Access to compare screen; can merge
• Online record creation and editing
• Reports on request
  • ISNIs – simple report or enhanced
  • Cluster movement report
  • Diagnostic reports
• Statistics and links
Adding a new record

Identifier: JNAM 99988889999

Name:
- fornames: Michel
- surname: Calame

Related organisation:
- University of Basel, Department of Physics

Relation:
- is related: -- select one --
- Relationship type: -- select one --

Relationship dates:
- begin:
- end:

PPN:

Language:

Gender:
- select one

Nationality:

Information:
- Source: Authors home page
- URI: http://calame.unibas.ch/

Source:
- Information: publicationslist.org
- URI: http://publicationslist.org/michel.calame

General note:

Save changes  Cancel
Adding your source to an existing record

<table>
<thead>
<tr>
<th>Identifiers and Names</th>
<th>Personal Details</th>
<th>Associations</th>
<th>Works</th>
<th>Class and Role</th>
<th>Sources and Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Save changes</strong></td>
<td><strong>Cancel</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **This record doesn't fit into this form**

<table>
<thead>
<tr>
<th>System</th>
<th>PPN</th>
<th>ISNI</th>
<th>ISNI status</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>133884252</td>
<td>0000000133884250</td>
<td>provisional</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>identifier type</th>
<th>identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BRTH</td>
<td>429076</td>
</tr>
</tbody>
</table>

- **Add Identifier**
  - Required when adding your source (JNAM)

- **Add Name**
  - Craig
  - Aaen-Stockdale
  - BRTH
  - Required when adding your source (JNAM)

- **Add Name variant**
These are all the same person. The second has an incorrect DOB = 1900
You can add a source note or general note to any database record, your code does not need to be present
Reporting errors

The general note will trigger an email to the ISNI Quality Team for attention
VIAF data in ISNI

• Held as source and sub source (because counted differently for determining assignment)

• Inherited VIAF data model

• VIAF diffuses ISNIs, e.g. to Wikipedia

• Special monthly update procedures

• ISNI base file; VIAF clustering accepted without re-clustering
Cluster movement

<table>
<thead>
<tr>
<th>Record ID</th>
<th>Action</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>BNF</td>
<td>11898958</td>
<td>add</td>
</tr>
<tr>
<td>LCjn 85175479</td>
<td>add</td>
<td>2009-03-03T12:03:36+00:00</td>
</tr>
<tr>
<td>DNBJ</td>
<td>39738274</td>
<td>add</td>
</tr>
<tr>
<td>NLII</td>
<td>DA01486338</td>
<td>add</td>
</tr>
<tr>
<td>NLII</td>
<td>DA01486338</td>
<td>delete</td>
</tr>
<tr>
<td>SUDOC</td>
<td>J026815664</td>
<td>add</td>
</tr>
<tr>
<td>LCjn 85175479</td>
<td>delete</td>
<td>2011-12-15T08:03:23+00:00</td>
</tr>
<tr>
<td>BNF</td>
<td>11898958</td>
<td>delete</td>
</tr>
<tr>
<td>BNF</td>
<td>11898958</td>
<td>add</td>
</tr>
<tr>
<td>LCjn 85175479</td>
<td>add</td>
<td>2012-05-22T07:07:49+00:00</td>
</tr>
<tr>
<td>NLII</td>
<td>atj000471441</td>
<td>add</td>
</tr>
<tr>
<td>NTAl</td>
<td>072465484</td>
<td>add</td>
</tr>
<tr>
<td>NLII</td>
<td>000471441</td>
<td>add</td>
</tr>
<tr>
<td>NLII</td>
<td>atj000471441</td>
<td>delete</td>
</tr>
<tr>
<td>ISNI</td>
<td>00000000080786322</td>
<td>add</td>
</tr>
<tr>
<td>LCjn 85175479</td>
<td>delete</td>
<td>2014-01-15T05:38:27+00:00</td>
</tr>
<tr>
<td>NLII</td>
<td>000471441</td>
<td>delete</td>
</tr>
<tr>
<td>NTAl</td>
<td>072465484</td>
<td>delete</td>
</tr>
<tr>
<td>NLPl</td>
<td>a27066502</td>
<td>add</td>
</tr>
<tr>
<td>LCjn 85175479</td>
<td>add</td>
<td>2014-05-14T22:19:01+00:00</td>
</tr>
<tr>
<td>NLII</td>
<td>000471441</td>
<td>add</td>
</tr>
<tr>
<td>NTAl</td>
<td>072465484</td>
<td>add</td>
</tr>
</tbody>
</table>

VIAF re-clusters every month and makes duplicate clusters where one source has duplicates. The result can be cluster movement. ISNI has been monitoring VIAF cluster movement & making recommendations.

Merges are valid cluster moves.
ISNI & VIAF XA Records

- VIAF includes XA records that act as “Police records”
- If an ISNI record has 2 VIAF Ids & an indication of a manual merge, the ISNI record gets XA status & will cause merges in VIAF
- If ISNI sends 2 records with the same name & an indication of manual split, it will cause a split in VIAF. A record that could go in either cluster will be admitted to neither
Mixed identities

Source 1  Source 2

Cluster Error

Source 1  Source 2

Source Error
Correction – VIAF Cluster Error

- ISNI marks its two records as verified & sends to VIAF
- These records are given the same status as XA records in VIAF clustering.
- No two XA records may occur in the same cluster
ISNI Quality Team*

- Samples data regularly
  - c. 2% VIAF clusters have mixed identities
  - Duplicate clusters are higher, nearer 5%
- Makes corrections at cluster level
  - Merges, splits, error notifications
  - Access to cataloguing client / macros
- Makes system recommendations
- Gives approval for single source assignment
- Responds to End User input
- Sends emails to sources for error correction (19 VIAF sources currently participating)

* ISNI Quality Team is expanding to include more libraries
Crowd sourcing

End users can send notifications via this yellow box.
Discussion and policy papers (internal)

Pseudonyms, Name changes and Maiden/Married names

An ISNI policy document

Contents

1. ISO Standard – Definition of ISNI
2. Change of Name – General Principles
3. ISNI data – VIAF/Rights Management Agencies
4. Pseudonyms
5. Married/Maiden Names
6. Gender Reassignment
7. Policies on Maiden/Married Names and Gender Reassignment
Discussion and policy papers (external)

OCLC Research Task Group on Organisations in ISNI

Karen Smith-Yoshimura
Grace Agnew
Christopher Brown
Kate Byrne
Matt Carruthers
Peter Fletcher
Janifer Gatenby
Stephen Hearn
Xiaoli Li
Marina Muilwijk
Chew Chiat Naun
Roderick Sadler
John Riemer
Jing Wang
Glen Wiley
Kayla Willey

OCLC Research (leader)
Rutgers University
Jisc (UK)
University of New South Wales
University of Michigan
UCLA
OCLC Leiden (ISNI Assignment Agent)
University of Minnesota
University of California, Davis
University of Utrecht
Cornell University
La Trobe University
UCLA
Johns Hopkins University
University of Miami
Brigham Young University

Examined 12 use cases; producing sample records for each use case
23 recommendations for the system, for the ISNI-IA, for users
Search guidelines for organisations
Outreach document

With input from Andrew MacEwan, British Library and Anila Angjeli, Bibliothèque nationale de France
Governance & growth

- 52 performer rights management organisations
- 89 text rights management organisations
- 229 music rights management organisations
- 60,000 member libraries
- 46 national libraries
- 46 national libraries
- COAPR
- Assignment Agency
- Data contributor
- ISNI Executive & Administration
Leverage, alignment & growth

Trade Sources

Text Rights

Music Rights

Encyclopaedias

Archives and Museums

Libraries

Researchers & Professional

Granting organisations

Professional Societies

Article databases

Theses databases
Alignment & growth: some areas of activity

- Organisation IDs for research management systems
- Organisation IDs for publishers & imprints
- Assignment in CIP workflows (BnF)
- ORCID to ISNI API interfaces
- Self-claim online ISNI request (data collection from Public interface)
- Batch-load International Performers Database (SCAPR – performers’ rights)
- Journal articles & theses
- Archives
- Downloadable LOD file
VIAF & ISNI alignment: considerations

• Sustainability of VIAF-ISNI updates
• National library direct loads to ISNI (BnF, KB,…)
• Business model for library members
• Development of ISNI Technical and Policy governance (including VIAF interoperability – a libraries’ issue)
• Development of ISNI systems and processes to accommodate Member recommendations & requests
• Aligning library participation with non-library participation: new service models for national libraries, research libraries, etc.
• **Opportunity for leveraging library authority data with different communities on the Web**
Thankyou